



# Patient Matching, Aggregation, and Linking (PMAL) Project

## FINAL REPORT

Office of the National Coordinator for Health Information Technology

August 2019

## TABLE OF CONTENTS

<b>Introduction .....</b>	<b>3</b>
<b>Target Areas .....</b>	<b>4</b>
<b>TARGET AREA 1: IMPROVEMENTS TO MATCHING ALGORITHMS.....</b>	<b>4</b>
Gold Standard & Algorithm Testing for Patient Matching Pilot .....	4
Patient Matching Algorithm Challenge.....	5
Patient Matching Test Harness.....	5
<b>TARGET AREA 2: IMPROVEMENTS TO DATA QUALITY.....</b>	<b>6</b>
Patient Demographic Data Quality Framework Pilot.....	7
<b>TARGET AREA 3: EXPANDED DATA SHARING .....</b>	<b>7</b>
Health Data Security Layer.....	8
Tools for the Creation of a Longitudinal Patient Record .....	10
Novel Approaches to Patient Data Sharing.....	11
<b>TARGET AREA 4: DATA STANDARDIZATION .....</b>	<b>12</b>
Record Completeness .....	12
Health Data Provenance Challenge .....	13
Provider Data .....	13
<b>Lessons Learned and Recommendations .....</b>	<b>14</b>
Health Care Provider Engagement .....	14
Open Source Software and Tools .....	15
Patient Data.....	15
Evolving Technology.....	15
Conclusion and Next Steps.....	15
<b>Appendix A – PMAL Project Products Summary .....</b>	<b>17</b>

## Introduction

The Department of Health and Human Services (HHS) Office of the Assistant Secretary for Planning and Evaluation (ASPE) engaged the Office of the National Coordinator for Health Information Technology (ONC) to study and advance patient matching, aggregation, and linking (PMAL, or the PMAL Project) through the Patient-Centered Outcomes Research (PCOR) Trust Fund. ASPE oversees federal health programs funded through the PCOR Trust Fund to build data capacity for research. PCOR studies are designed to produce new scientific evidence that informs and supports the health care decisions of patients, families, and their health care providers. PCOR studies examine the effectiveness of prevention and treatment options while taking into consideration the preferences, values, and questions that are important to patients when they make health care choices.

The PMAL Project<sup>1</sup> was intended to study and advance methods of accurately linking and aggregating patient records from various sources to increase the volume and quality of data available to support research. Accurate patient matching and data aggregation offers additional benefits, such as increased patient safety and improved health and well-being.

The PMAL Project began in June 2015 and ended September 2018. The project identified and tested standards applicable to linking patient data across multiple types of data sets (e.g., research, clinical, and claims) to support research data infrastructure. The project included pilot projects and other activities, such as prize challenges, to address data quality, help improve patient match rates, and promote data standardization to advance interoperability. The project also explored advancements in technology that had the potential to support patient-mediated data exchange, improved patient matching, and improved data security.

The PMAL Project addressed the challenges of patient matching and linking in four target areas:

- **Improvements to Matching Algorithms** – The project sought to increase transparency in the performance of existing patient matching algorithms, spur the adoption of performance metrics by patient data matching algorithm developers, and enable researchers to more accurately link patient data from different sources.
- **Improvements to Data Quality** – The project examined ways to improve the quality and accuracy of patient data captured in clinical systems, which has a large impact on the performance of patient matching algorithms. Higher quality patient data leads to better match rates and fewer false positives.
- **Expanded Data Sharing** – The project considered approaches to security and data access mechanisms. Access to patient data from different sources and the ability to share data is a prerequisite to patient matching and linking, which require a trusted security layer and standard mechanisms for data access.

---

<sup>1</sup> See, e.g., PMAL Project at <https://www.healthit.gov/topic/scientific-initiatives/pcor/patient-matching-aggregating-and-linking-pmal>.

- **Data Standardization** – Patient matching algorithms depend on a set of data elements such as patient name, address, and other information. Standardizing the content and form of these data elements is a key requirement for the development of algorithms that can match data from different sources.

Over the course of the project, ONC produced multiple deliverables (as shown in Table 1, Appendix A) to facilitate PCOR, including a variety of activities intended to engage industry innovators. This work also established a foundation for further testing and development of patient matching algorithms. In some cases, a single deliverable addressed multiple target areas. In those cases, the deliverable is described in detail in the most relevant target area and then, to avoid duplication, only referenced in other target areas.

## Target Areas

### TARGET AREA 1: IMPROVEMENTS TO MATCHING ALGORITHMS

The PMAL Project focused on improving the algorithms used for patient matching, increasing transparency regarding the performance of existing patient matching algorithms, and spurring the adoption of performance metrics by patient data matching algorithm vendors. PCOR benefits from improved algorithms that match and link patient data because this contributes to accurately aggregated data across multiple institutions. Accurately matched data improves researchers' ability to track patient outcomes and increases their confidence that they are viewing correct, complete data.

To advance improvements to patient matching algorithms under the PMAL Project, ONC:

- Developed a gold standard data set and used it to test a widely adopted patient matching algorithm in the Gold Standard & Algorithm Testing (GSAT) for Patient Matching Pilot
- Launched the Patient Matching Algorithm Challenge to increase the availability of information regarding the performance of patient matching algorithms
- Implemented an open source, patient matching test harness and the Identity Matching Adjudicator Collector (IMAC)<sup>2</sup> for use in the GSAT pilot and the Patient Matching Algorithm Challenge

### Gold Standard & Algorithm Testing for Patient Matching Pilot

The Gold Standard & Algorithm Testing (GSAT) Pilot Project was a collaboration between ONC, OCHIN (a health IT support and service provider for safety-net clinics), and the Kaiser Permanente Center for Health Research's (CHR) Data Coordinating Center. The Patient Matching Community of Practice, a separate effort led by ONC and the Healthcare Information and Management Systems Society (HIMSS), also contributed to the GSAT Project by providing feedback as the project began.<sup>3</sup> This project pilot tested the performance of a widely adopted patient matching algorithm to quantify patient matching accuracy.

---

<sup>2</sup> <https://github.com/mitre/IMAC>

<sup>3</sup> <https://www.healthit.gov/sites/default/files/ptmatchwhitepaper.pdf>

In the course of the project, a gold standard data set was created to compare patient record pairs identified as matches by matching algorithms to the known duplicates in the gold standard data set, and to then compute evaluation metrics. The project tested the performance of a widely-adopted electronic health record (EHR) patient matching algorithm.

Study results indicated that other health system providers should be encouraged to test the performance of the patient matching algorithms that are available to them. Algorithm developers should also test the parameters, settings, and design of their patient matching algorithms against carefully constructed test data that are representative of the base data set demographics and cultural origin of names. The results from this pilot test also informed other activities and deliverables of the PMAL Project. The GSAT Project and similar studies may promote innovation in patient matching and open a dialogue about patient matching algorithms among vendors, developers, and health care providers.

### **Patient Matching Algorithm Challenge**

ONC hosted a Patient Matching Algorithm Challenge<sup>4</sup> to bring about “greater transparency and data regarding the performance of existing patient matching algorithms” and to “spur the adoption of existing, widely accepted, algorithm performance metrics by patient data matching algorithm developers.”

The challenge had participants link records using demographic data. The participants applied the algorithms they developed to a synthetic data set that was provided to all participants.<sup>5</sup> ONC evaluated and scored the matches made by their algorithms against a master (or answer) key specifically developed for the synthetic data set. The master key consisted of the set of true match pairs created and verified through manual review within the larger synthetic data set. The performance of participants’ algorithms was judged using the well-accepted, information retrieval measures of precision, recall and F-score.

At the completion of the challenge, top-performing teams were interviewed to determine which data elements were most useful to improving performance. Due to popular demand, ONC hosted the environment for another year past the completion of the challenge to give others the opportunity to test their algorithms.<sup>6</sup> This demonstrates how a challenge competition can be used to increase awareness and participation in solving complex issues in patient matching.

### **Patient Matching Test Harness**

Researchers or other entities may have a need to compare the performance of two different matching algorithms. Typically, matching algorithms have their own data input and output formats along with their own methods for starting matching jobs.

---

<sup>4</sup> <https://www.patientmatchingchallenge.com>

<sup>5</sup> The PMC dataset is currently available at <https://www.patientmatchingchallenge.com>.

<sup>6</sup> The environment was built using a server less architecture to minimize cost and enable it to run for a long duration at low expense. As of May 2019, the environment is still live.

The goal of the open source, patient matching test harness software developed under the PMAL Project was to provide an environment where it is easy to run and tune multiple patient matching algorithms and view performance metrics.<sup>7</sup> The test harness defines a set of Health Level 7 International® (HL7®) Fast Healthcare Interoperability Resources® (FHIR®)-based methods for matching algorithms to start a matching process, access records to match, and report results of the matching process. FHIR is a standards-based framework for building interoperable health information exchange specifications.<sup>8</sup> Once an algorithm has been integrated with this FHIR-based interface, users of the test harness can easily run matching processes against the user's different data sets or with different matching thresholds to determine differences in performance metrics.

A simplistic patient matching algorithm was included as part of the patient matching test harness package.<sup>9</sup> This algorithm demonstrates how to use the test harness and how algorithms can integrate with the FHIR-based interface.

The test harness was also developed to support the GSAT Project. Because the harness could run matching algorithms against different data sets, it could run the same algorithm against a patient set from month to month to collect metrics of interest, such as database duplicate rate. ONC also developed the Identity Matching Adjudication Collector tool to assist with the verification of matches in data sets. Although the IMAC tool was created for the GSAT Project specifically, it is now publicly available for use across research, academic, and commercial communities.

## TARGET AREA 2: IMPROVEMENTS TO DATA QUALITY

The second target area of the PMAL Project was the improvement of patient demographic and clinical data quality by providing guidance to support the collection of quality demographic data from patients as well as maintaining data quality throughout its travels in the health care system. The quality and accuracy of patient data captured in clinical settings has a significant impact on the performance of patient matching algorithms. Higher quality patient data leads to better match rates and fewer false positives, which benefits all users of the data, including clinicians, researchers, and patients. This target area is aligned with the recent Patient Matching request for information (RFI) in ONC's 21<sup>st</sup> Century Cures Act proposed rule published in February 2019.<sup>10</sup> The RFI sought input on the potential effect that data collection standards may have on the quality of health data that is captured and stored, and the impact that such standards may have on accurate patient matching.

To advance the improvement of data quality, ONC undertook the following actions under the PMAL Project:

- Pilot tested an implementation of the Patient Data Demographic Quality (PDDQ) Framework
- Developed the Record Completeness Service (Scorecard) (discussed in Target Area 4 on page 10)

---

<sup>7</sup> The record matching interface can be found on Github at <https://mitre.github.io/test-harness-interface/>.

<sup>8</sup> See, e.g., [https://www.hl7.org/implement/standards/product\\_brief.cfm?product\\_id=491](https://www.hl7.org/implement/standards/product_brief.cfm?product_id=491).

<sup>9</sup> See, e.g., <https://github.com/mitre/ptmatchadapter/tree/master/examples/ptmatchadapter-fril>.

<sup>10</sup> ONC 21<sup>st</sup> Century Cures Act proposed rule, February 2019:  
<https://www.federalregister.gov/documents/2019/03/04/2019-02224/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification>



- Launched the “Oh the Places Data Goes: Health Data Provenance Challenge” (discussed in Target Area 4 on page 10)

### Patient Demographic Data Quality Framework Pilot

High-quality patient demographic data is fundamental to accurate patient identification and matching. Consequently, accurate patient identification and matching is pivotal to interoperability, quality of care, patient safety, and research, such as PCOR. With the adoption of rapid developments in health information technology (IT) and advancements in EHRs, the scale of growth of the data captured, stored, and exchanged continues to increase. Furthermore, the constant growth in patient populations and their diversity, along with the increasing complexity of health care networks, means that health care staff must have the capabilities and best practices to capture high-quality data on the front line.

ONC tested the application of the Patient Demographic Data Quality Framework in a pilot implementation.<sup>11</sup> The PDDQ Framework is a health care-specific demographic data management model that is based on the Capability Maturity Model Integration (CMMI) Institute’s Data Maturity Model<sup>SM</sup>. The PDDQ Framework provides guidance on standardizing policies, procedures, and practices both locally and across an entire health care system. The standards and processes outlined in the framework offer guidance to health care organizations on improving the quality of patient demographic data by preventing the creation of duplicate patient records and better managing existing duplicate records.

From December 2016 through June 2017, ONC collaborated with OCHIN and the Kaiser Permanente Center for Health Research to test the PDDQ Framework at participating OCHIN-affiliated Community Health Centers. The results of the pilot study suggest there is a potential for improvements in data quality with a modest investment. As a result of the pilot, ONC captured lessons learned that organizations may consider in the implementation of the PDDQ Framework and encouraged organizations to participate in the development of guidance resources that improve the quality of demographic data and patient matching. ONC also identified the opportunity for further study to better establish the correlation between application of the PDDQ Framework and its impact on metrics that track duplicate records.

### TARGET AREA 3: EXPANDED DATA SHARING

prerequisite to patient matching and linking is access to patient data from different sources. Even with advancements in interoperability, there are barriers to sharing data that can hinder PCOR. These barriers include a lack of trust between parties who might otherwise share data and a dearth of technology that can effectively aggregate data gathered from disparate sources. Under the PMAL Project, ONC addressed barriers to data sharing through the following tasks:

- Developed a trusted security layer, including specifications and tools, to enable patient control regarding how their data is shared
- Developed tools and reference implementations to facilitate data sharing and assembly of a complete longitudinal patient record from multiple disparate sources

---

<sup>11</sup> See, e.g., the PDDQ Framework Pilot Project at <https://www.healthit.gov/topic/scientific-initiatives/pcor/patient-matching-aggregating-and-linking-pmal>.

- Launched two public challenges focused on novel approaches to patient data sharing

## Health Data Security Layer

ONC, in collaboration with the OpenID Foundation,<sup>12</sup> established the Health Relationship Trust (HEART) Working Group.<sup>13</sup> The HEART Working Group was intended to gather representatives from the many different health-related technical communities working in such areas as patient authentication, authorization, and consent, to collaborate on normative specifications and profiles that meet their shared goals. The HEART Working Group developed a set of privacy and security specifications (known as HEART specifications) that build on current industry practices using existing, broadly accepted technologies. These included FHIR, OAuth,<sup>14</sup> OpenID Connect,<sup>15</sup> and User-Managed Access (UMA).<sup>16</sup> The following specifications build on industry practice and existing state-of-the-art security and add new components for management of patient-mediated interoperability of health information:

- **Health Relationship Trust Profile for OAuth 2.0**<sup>17</sup> – This specification defines how to use OAuth 2.0 to securely delegate access to health data.
- **Health Relationship Trust Profile for FHIR OAuth 2.0 Scopes**<sup>18</sup> – This specification defines how to use OAuth 2.0 scopes with the FHIR protocol to securely delegate access to health data.
- **Health Relationship Trust Profile for User-Managed Access 2.0**<sup>19</sup> – This specification defines how to use UMA in the context of managing access to health data.

---

<sup>12</sup> <https://openid.net/foundation/>

<sup>13</sup> <https://openid.net/wg/heart/>

<sup>14</sup> See, e.g., OAuth 2.0, available at <https://oauth.net> (an “open protocol to allow secure authorization in a simple and standard method from web, mobile, and desktop applications.”). This profile allows third-party applications to access web services on a limited basis.

<sup>15</sup> See, e.g., OpenID, available at <https://openid.net/connect/>. Used with OAuth 2.0, this profile enables clients to verify the identity of an end-user and to provide information about authenticated sessions and end users.

<sup>16</sup> See, e.g., UMA Home, available at <https://kantarainitiative.org/confluence/display/uma/Home>. UMA is a protocol based on OAuth that is “designed to give a web user a unified control point for authorizing who and what can get access to their online personal data...content...and services...no matter where all those things live on the web.” *Id.*

<sup>17</sup> Richer, J., Ed., Health Relationship Trust Profile for OAuth 2.0 (openid-heart-oauth2-1\_0) (July 8, 2018), available at [https://openid.net/specs/openid-heart-oauth2-1\\_0.html](https://openid.net/specs/openid-heart-oauth2-1_0.html).

<sup>18</sup> Richer, J., Ed., Health Relationship Trust Profile for Fast Healthcare Interoperability Resources (FHIR) OAuth 2.0 Scopes (openid-heart-fhir-oauth2) (July 8, 2018), available at [https://openid.net/specs/openid-heart-fhir-oauth2-1\\_0.html](https://openid.net/specs/openid-heart-fhir-oauth2-1_0.html).

<sup>19</sup> Richer, J., Ed., Health Relationship Trust Profile for User-Managed Access 2.0 (openid-heart-uma-2\_0) (July 8, 2018), available at [https://openid.net/specs/openid-heart-uma2-1\\_0.html](https://openid.net/specs/openid-heart-uma2-1_0.html).



- **Health Relationship Trust Profile for FHIR UMA 2 Resources**<sup>20</sup> – This specification defines how UMA and FHIR are combined to allow a patient to securely delegate access to their health data using the FHIR protocol.

Development and testing of the HEART specifications also produced contributions that refined these underlying technologies. For example, the HEART working group generated 28 change requests to the UMA specification, ensuring health data use cases were adequately covered by the specification. The HEART working group also worked with Argonaut,<sup>21</sup> a private sector initiative that seeks to advance industry adoption of interoperability standards, to ensure compatibility between the Argonaut and HEART specifications to make it straightforward for Argonaut to add user-managed access in the future.<sup>22</sup>

In addition to developing the HEART specifications, open-source reference implementations of the HEART specifications were also developed. These reference implementations:

- Ensured the specifications were of sufficient quality and clarity to be implementable
- Provided open source implementations for use in implementation and interoperability testing
- Provided accessible resources as potential foundations for commercial and open source implementations

The following open source reference implementations were developed for the HEART specifications:

- The OpenID Connect and OAuth 2.0 specifications that formed the foundation of the HEART specifications<sup>23</sup>
- A “toolkit” to enable vendors to create HEART-compliant clients and services.<sup>24</sup>

ONC launched two events to promote the HEART specifications and engage the broader health IT community in actively incorporating HEART or strategizing ways to improve it:

### **Move Health Data Forward Challenge<sup>25</sup>**

---

<sup>20</sup> Richer, J., Health Relationship Trust Profile for Fast Healthcare Interoperability Resources (FHIR) UMA 2 Resources (openid-heart-fhir-uma-2\_0) (July 8, 2018), available at [https://openid.net/specs/openid-heart-fhir-uma2-1\\_0.html](https://openid.net/specs/openid-heart-fhir-uma2-1_0.html).

<sup>21</sup> See, e.g., HL7 International, Welcome to the Argonaut Project, available at [http://argonautwiki.hl7.org/index.php?title=Main\\_Page](http://argonautwiki.hl7.org/index.php?title=Main_Page).

<sup>22</sup> Argonaut App-Authorization Profile Alignment with HEART Profiles (December 30, 2015), available at <http://argonautwiki.hl7.org/images/b/b7/%282015Dec30%29HEARTcomparisonV3.pdf>.

<sup>23</sup> See, e.g., <https://github.com/mitreid-connect/>, which contains a selection of tools such as an OpenID Connect reference implementation in Java and a web application demonstrating the use of the OpenID Connect client code and configuration.

<sup>24</sup> See, e.g., Go HEART at <https://github.com/mitre/heart>, which is a library that provides tools for go-based clients to enable them to conform to the HEART Working Group profiles. In addition, an existing FHIR server, Intervention Engine, was leveraged that could support the HEART specifications.

<sup>25</sup> HHS, Move Health Data Forward Challenge, available at <https://www.challenge.gov/challenge/move-health-data-forward-challenge/>.

This challenge invited participants to create an API solution that used the HEART specifications to allow individuals to securely authorize the movement of their health data to destinations they chose. This could include sharing for research purposes. The challenge began in November 2016 and ended in January 2017. Winners were awarded in each of three distinct phases (Phase 1 focused on proposals; Phase 2 required a prototype and pilot testing; and Phase 3 required a test of the solution on real-life situations and which demonstrated scalability).

### **Proving the Potential: A Health Data and Standards Code-A-Thon<sup>26</sup>**

This challenge invited teams to select one of three use cases for which to develop ways to seamlessly: (1) incorporate electronic clinical quality measures and clinical decision support, (2) ensure the security and privacy of aggregated patient data, or (3) discover researchers, providers, and/or services. This challenge helped to enhance interoperability and demonstrate ways to configure systems more easily and securely to share data for research as well as connect patients with researchers. In addition, it emphasized using open source tools and services to build more advanced services and showcase their innovations. The challenge required the use of one or more named tools, including the HEART profiles and the patient matching test harness described elsewhere in this report. Winners were announced for first, second, and third place.

### **Tools for the Creation of a Longitudinal Patient Record**

In conjunction with data standardization, research analysis benefits from having tools to more easily aggregate and track patient data. ONC therefore undertook several smaller projects under the PMAL Project to develop tools that would facilitate patient data sharing and the creation of a longitudinal patient record from multiple disparate sources. These tools can be found in the FHIR section along with other ONC Testing Tools and Utilities on the ONC Tech Lab Website.<sup>27</sup>

ONC developed a set of FHIR-based tools to illustrate the use of FHIR for following use cases:

- Sharing of data via a FHIR Server<sup>28</sup>
- Calculation of electronic clinical quality measures on FHIR-based patient record<sup>29</sup>
- Development of an API to make FHIR data accessible to tools built for older standards (HL7 QRDA Category I)<sup>30</sup>
- Merging of two FHIR-based patient records using an interactive Record Merge Tool<sup>31</sup>

---

<sup>26</sup> CCC Innovation Center, Proving the Potential: A Health Data and Standards Code-A-Thon, available at <https://www.cccinnovationcenter.com/challenges/proving-the-potential-a-health-data-and-standards-code-a-thon/>.

<sup>27</sup> [https://www.healthit.gov/techlab/testing\\_and\\_utilities.html](https://www.healthit.gov/techlab/testing_and_utilities.html)

<sup>28</sup> Intervention Engine, *supra* note 22.

<sup>29</sup> <https://github.com/mitre/ecqm>

<sup>30</sup> <https://github.com/mitre/fhir-patient-api>

<sup>31</sup> Record Merge Tool, <https://github.com/mitre/ptmerge>

These tools, which are available on the ONC website,<sup>32</sup> were intended to make FHIR more accessible for developers with less expertise in health care and to demonstrate the versatility of FHIR-based access to patient data.

A claims data importer tool was also developed to combine clinical and claims information to support accurate calculation of clinical quality measures. Many health information exchange organizations receive both clinical and claims data for the same individuals and can therefore benefit from this capability.

The claims data importer tool demonstrated the potential to merge claims data with FHIR-based clinical data. Combined with the tools for merging patient records and calculating electronic clinical quality measures, the claims data importer tool provided an example for health information exchange organizations that lacked such capabilities. The claims data importer tool supports ASC X12 837 Professional (837P) transactions, which is the standard used to detail claims submitted by individual providers for outpatient encounters. Because ASC X12, a standards organization, has intellectual property restrictions that inhibit publishing related code as open source, the software for the claims data importer tool is available from ONC upon request.<sup>33</sup>

### Novel Approaches to Patient Data Sharing

ONC hosted two events to explore the potential for applying distributed ledger technology, often referred to as blockchain, to health IT challenges. Blockchain technology solutions are being considered by the health IT industry for use cases that require a durable and distributed record of transactions. This technology is intended to provide a trustworthy means to facilitate coordination between domains and especially where information is asymmetric and moves over time in a potentially untrustworthy environment. The goal of these events was to explore new and novel approaches to sharing patient data.

The first event, the “Use of Blockchain in Health IT and Health-Related Research,” was an ideation challenge that solicited white papers on the topic of blockchain technology and its potential use in health IT to address privacy, security, and scalability challenges of managing EHRs and resources.<sup>34</sup> ONC accepted submissions to the challenge between July and August 2016. Fifteen whitepapers were selected as winning entries, and eight winners were invited to present their papers at an industry-wide workshop co-hosted with the National Institute of Standards and Technology (NIST). Notably, one of the finalists in the “Oh the Places Data Goes Challenge” (described in Target Area 4 on page 10) was a blockchain-based solution, which demonstrates the intersection of these technologies and their importance to innovation.

ONC hosted a code-a-thon in March 2017 to challenge contestants to create projects that incorporated distributed ledger technology. The Blockchain in Healthcare Code-A-Thon required contestants to address one of the following topics: identity management and application programming interfaces (API); metadata tagging and policy expression to manage access and provide auditing and provenance information, or data

---

<sup>32</sup> [https://www.healthit.gov/techlab/testing\\_and\\_utilities.html](https://www.healthit.gov/techlab/testing_and_utilities.html)

<sup>33</sup> To access the claims data importer software, a request may be made to ONC: [ONC.Request@hhs.gov](mailto:ONC.Request@hhs.gov).

<sup>34</sup> ONC, Announcing the Blockchain Challenge, “Use of Blockchain in Health IT and Health-Related Research,” <https://www.healthit.gov/topic/grants-contracts/announcing-blockchain-challenge>.

aggregation and linkage for shared patient data.<sup>35</sup> Contestants had 24 hours to develop their solutions using open source distributed ledger technology and health-specific standards. Ten teams completed and submitted final projects.<sup>36</sup> The first place project provided a blockchain solution for a decentralized personal health record that gives control of health data to the patient and allows health information to flow across boundaries, relying on the patient to be a trusted authority between disparate systems.<sup>37</sup>

## TARGET AREA 4: DATA STANDARDIZATION

Patient matching algorithms depend on a set of data elements such as patient name, address, and other information. Standardizing the content and form of these data elements is a key requirement for the development of algorithms that can match data from various sources. This area of the PMAL Project focused on the standardization of data elements used by patient matching algorithms, with the goal of eventually providing patient matching algorithms with a consistent set of inputs regardless of the data sources. Additionally, this study is timely and aligned with the Patient Matching RFI in ONC's 21<sup>st</sup> Century Cures Act proposed rule,<sup>38</sup> which sought input on a required minimum set of elements that need to be collected and exchanged, as well as input on standardized metrics for the performance evaluation of available patient matching algorithms.

### Record Completeness

Complete patient records are important for both research and patient matching. Patient matching algorithms rely on a set of data elements within each patient record. Therefore, having a core set of data elements is critical for matching. In addition, an incomplete record can skew research. For example, when a patient record does not report allergies, does that mean the patient does not have any allergies or that allergies have not been recorded? This can be resolved by making a “no known allergies” entry in the patient record.

ONC developed the Patient Record Scorecard application to compute a completeness score for a FHIR-based patient record.<sup>39</sup> The score is based on which of the data fields defined by FHIR are actually present in the record, including negative assertions such as “no known allergies.” The application builds on SMART on FHIR,<sup>40</sup> a platform that facilitates the development of applications that can securely run across the health care system and allows its use with any SMART on FHIR-compliant source of patient records such

---

<sup>35</sup> Blockchain in Healthcare Code-A-Thon at <https://www.cccinnovationcenter.com/challenges/blockchain-in-healthcare-code-a-thon/>; see also <https://oncprojecttracking.healthit.gov/wiki/display/TechLab/The+Blockchain+in+Healthcare+Code-A-Thon>.

<sup>36</sup> Descriptions of the completed projects are available at <https://oncprojecttracking.healthit.gov/wiki/display/TechLab/Code-A-Thon+Projects>.

<sup>37</sup> *Id.*

<sup>38</sup> ONC 21<sup>st</sup> Century Cures Act proposed rule, February 2019: <https://www.federalregister.gov/documents/2019/03/04/2019-02224/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification>

<sup>39</sup> See <https://github.com/mitre/scorecard> app.

<sup>40</sup> See <https://smarthealthit.org>.

as an EHR system. The Patient Record Scorecard application takes a FHIR patient record as input and returns a score as output.

The scoring algorithm is based on the C-CDA Scorecard<sup>41</sup> and provides users with information on the quality of the evaluated EHR. ONC successfully tested the Record Completeness application against Cerner Millennium and Epic EHR systems, demonstrating its utility against patient records captured by different EHR vendor products. Additionally, the code base from the Patient Record Scorecard application was used to create Inferno, which is an ONC FHIR testing tool.<sup>42</sup>

## Health Data Provenance Challenge

Data provenance is the information about the entities, activities, and people involved in production of data. Data provenance information can identify the original creator of health information, when and where the information was created, what has changed in the information, and why changes were made. Data provenance can give data recipients confidence in the data's authenticity, trustworthiness, and reliability. This information can be invaluable to researchers where the rigor of the study can depend on confidence in these data traits. In patient matching and linking, data provenance is fundamental to patient safety as well as to assuring secure health information exchange.

To advance work in the area of health data provenance, ONC hosted the “Oh the Places Data Goes: Health Data Provenance Challenge.”<sup>43</sup> This challenge involved two issues. One was identifying real-world provenance problems, understanding why they are important to solve, and providing participants an opportunity to develop practicable and executable plans for innovative solutions. The other was to demonstrate viable solutions to provenance problems in ways that demonstrated high technological merit, scalability, feasibility of implementation, and impact.

There were two winning solutions. One solution pilot tested the use of its partner's provider application to surface provenance information and help providers find aggregated data from various sources using FHIR. The other solution created a software toolkit enabling health information systems of any size to integrate provenance into their data-flow without disrupting existing practices or data repository requirements. The winning participants demonstrated innovative, standardized solutions that held the potential to improve health data provenance.

## Provider Data

Correctly identifying a provider across data sets can assist the process of linking and aggregation of patient data. Existing provider directories often contain inaccurate information or are inconsistent across various patient-related data sources. ONC, in collaboration with the Federal Health Architecture (FHA), sponsored a Provider Directory Workshop to gather public and private stakeholders to review challenges, share successes, and generate new ideas on provider directory standards and solutions to further interoperability.

---

<sup>41</sup> <https://sitenv.org/ccda-smart-scorecard/>

<sup>42</sup> <https://inferno.healthit.gov/>

<sup>43</sup> <https://www.cccinnovationcenter.com/challenges/provenance-challenge/>

The workshop resulted in the following activities:

- The development of an open API for the National Plan and Provider Enumeration System (NPPES).<sup>44</sup> NPPES manages the assignment of standard unique identifiers for health care providers and health plans. The open API for NPPES allows systems to query the NPPES for provider information, ensuring that it is captured uniformly in systems that use provider data and linked to patient records such as EHR systems.<sup>45</sup> ONC worked collaboratively with the Centers for Medicare & Medicaid Services to implement the API for NPPES on the CMS Web site and develop a coordination and outreach plan to promote its adoption and use by any stakeholder needing to obtain information on health care providers and health plans.
- ONC created a set of provider data tools for parsing, creating, and validating United States-based health provider data.<sup>46</sup> These tools could be integrated with health information exchange organizations or research systems to simplify the use of this data in their systems.
- ONC developed an OAuth client and server libraries to manage authentication and authorization for users of the open API for NPPES.<sup>47</sup> This ensures that only authorized users can access the information and that CMS can track who accesses what information.

These activities generated high interest. Their results have been reused in other PCOR activities such as the pilot systems developed for CMS's Blue Button 2.0 initiative, part of the MyHealthEData initiative, which is intended to empower patients by giving them better control of their health care information.<sup>48</sup> Researchers and providers can use the outputs of these activities to obtain curated provider information, incorporate it into their systems, and enhance patient matching and linking capabilities.

Underscoring the perceived utility of the NPPES open API, HL7 is further developing it as a FHIR standard as the Validated Healthcare Directory Implementation Guide.<sup>49</sup>

## Lessons Learned and Recommendations

The PMAL Project revealed several issues of potential relevance to future studies on or work to improve patient matching.

### HEALTH CARE PROVIDER ENGAGEMENT

It was difficult to find health organizations willing to participate in patient matching projects and discussions because of a reluctance to openly discuss challenges with patient matching. In some cases,

---

<sup>44</sup> <https://nppes.cms.hhs.gov/#/>

<sup>45</sup> <https://npiregistry.cms.hhs.gov/registry/help-api>

<sup>46</sup> <https://github.com/HHSIDEALab/provider-data-tools>

<sup>47</sup> [https://github.com/TransparentHealth/hhs\\_oauth\\_client](https://github.com/TransparentHealth/hhs_oauth_client) and  
[https://github.com/TransparentHealth/hhs\\_oauth\\_server](https://github.com/TransparentHealth/hhs_oauth_server)

<sup>48</sup> <https://bluebutton.cms.gov>. Blue Button 2.0 is an API that allows Medicare beneficiaries to connect their Medicare claims data to trusted parties, including research programs.

<sup>49</sup> <http://hl7.org/fhir/uv/vhdir/2018Jan/index.html>



providers lacked resources to participate. These resource constraints, including lack of staff, also caused project delays or necessitated scaling back original project goals.

## OPEN SOURCE SOFTWARE AND TOOLS

Most, if not all, patient matching systems require proprietary integration. This presented a challenge for some tasks and goals of this project. Achieving standardization among systems will be necessary to facilitate patient data matching across the health care and research ecosystems.

The PMAL project sought to publish tools and resources under open source licensing when appropriate. It promoted the use of community-driven standards and fostered open discussions about the most effective and innovative ways of addressing patient matching, aggregation, and linking issues in support of building data capacity for PCOR.

## PATIENT DATA

Handling sensitive data such as patient data raises unique issues. Several project and development activities, such as the prize challenges, relied on synthetic data rather than patient data. For example, the Patient Match Challenge successfully used a synthetic data set to achieve its goals; however, the team still felt the need for more realistic, synthetic patient data that reflects real world situations. For example, matching algorithms are designed to reject invalid input. In practice, when patient identifiers are transformed or entered in an invalid way, the matchers will ignore or overlook them, resulting in inaccurate matches. The improvement of patient matching algorithms will require additional work on multiple fronts.

Of note, implementing the HEART specifications proved more complicated than expected due to the HEART specifications' technology choices and the strict security requirements necessitated by the specifications' health data focus. Recent developments such as the National Institutes for Health (NIH) *All of Us* Research Program<sup>50</sup> have increased focus on patient-controlled sharing of health data. The features and benefits of the HEART specifications may be leveraged for this and similar programs.

## EVOLVING TECHNOLOGY

Some of the projects were more challenging because of the continuously evolving state of technology. At times, the projects or challenges presented were innovative (e.g., the blockchain challenge) and their application in health IT had not yet been documented. Although this can be helpful in ushering industry toward innovative solutions, it also may mean, as was witnessed in this project, that there are few examples to emulate. This was the case in the development of the HEART specifications, which focused on patient-controlled sharing of health data. The health IT developer community has been focused mostly on enabling internal or provider-to-provider data sharing use cases. Although this type of issue is inherent to innovation, additional time for soliciting feedback and advanced discussions could have helped potential participants feel better prepared.

## CONCLUSION AND NEXT STEPS

ONC identified and tested standards that are applicable to linking patient data across multiple types of data sets (e.g., research, clinical, and claims) to support PCOR data infrastructure. Through multiple projects and activities, ONC developed a comprehensive set of tools and resources that support the

---

<sup>50</sup> <https://allofus.nih.gov/>

improvement of data quality, patient match rates, and promote data standardization to advance interoperability. ONC also explored technology advancements with the potential to support patient-mediated data exchange, improved patient matching, and improved security. The tools and base code developed by this project have been used by academic and private organizations to develop their own patient matching tools that suit their needs or test their patient matching systems.

In January 2019, the GAO published a report on patient matching, *Approaches and Challenges to Electronically Matching Patients' Records across Providers*,<sup>51</sup> as directed by the 21<sup>st</sup> Century Cures Act. In this report, GAO describes (1) stakeholders' patient record matching approaches and related challenges; and (2) efforts to improve patient record matching identified by stakeholders. Following this publication, the ONC published a Request for Information on patient matching as part of the ONC 21<sup>st</sup> Century Cures Act proposed rule published to the *Federal Register* in February 2019. In the RFI, ONC sought comment on additional opportunities that may exist in the patient matching space and ways that ONC can lead and contribute to coordination efforts with respect to patient matching. ONC noted particular interest in how patient matching can facilitate improved patient safety, better care coordination, the quality of care and advanced interoperability. ONC intends to review the responses to the RFI in concert with the GAO report to further inform continued efforts to advance patient matching building on the lessons learned from the PMAL Project. These initiatives and projects served to increase the capability to accurately link patient data from different sources and increase the volume of data to support PCOR research, as well as increase patient safety and improve health outcomes and well-being.

---

<sup>51</sup> U.S. Government Accountability Office, *Approaches and Challenges to Electronically Matching Patients' Records across Providers*, GAO-19-197, <https://www.gao.gov/assets/700/696426.pdf>

## Appendix A – PMAL Project Products Summary

Table 1 presents a summary of products produced and tools delivered by Target Area for the PMAL Project. The products produced during this task align to the following ASPE Intra-agency Agreement (IAA) objectives:

- **Objective 1:** Improve data quality, standardize data elements and improve algorithm match rates.
- **Objective 2:** Create an open source visual tool for patient matching and aggregation.
- **Objective 3:** Define the security layer necessary to protect application programming interfaces (API) and facilitate the aggregation of data from multiple data sources.
- **Objective 4:** Include Clinical Data Research Networks (CDRN) and their nodes in the piloting and testing of standards and services.
- **Objective 5:** Integrate with the National Plan and Provider Enumeration System (NPPES).

**Table 1: Products Summary for PMAL Project**

Target Area	Goals	Products / Tools Developed	Relevant Activity / Application	Alignment to ASPE IAA Objective
<b>1: Improvements to Matching Algorithms</b>				
Gold Standard & Algorithm Testing	<ul style="list-style-type: none"> <li>• Build a data set against which to effectively test performance of patient matching algorithms</li> </ul>	<ul style="list-style-type: none"> <li>• Gold Standard &amp; Algorithm Testing Whitepaper</li> </ul>	<ul style="list-style-type: none"> <li>• Gold Standard &amp; Algorithm Testing Pilot</li> </ul>	1, 4
Patient Matching Algorithm Challenge	<ul style="list-style-type: none"> <li>• Increase transparency and data on performance of existing patient matching algorithms, and</li> <li>• Spur adoption of performance metrics for algorithm vendors</li> </ul>	<ul style="list-style-type: none"> <li>• Challenge artifacts detailing the challenge and lessons learned</li> </ul>	<ul style="list-style-type: none"> <li>• Patient Matching Algorithm Challenge</li> </ul>	1

Target Area	Goals	Products / Tools Developed	Relevant Activity / Application	Alignment to ASPE IAA Objective
Patient Matching Test Harness	<ul style="list-style-type: none"> <li>Provide an environment in which it is easy to run and tune multiple patient matching algorithms and view performance metrics</li> </ul>	<ul style="list-style-type: none"> <li>Identity Matching Adjudicator Collector (IMAC) Tool</li> <li>Patient Matching Test Harness</li> </ul>	<ul style="list-style-type: none"> <li>Patient Matching Algorithm Challenge (used both the IMAC and the Patient Matching Test Harness)</li> </ul>	1, 2
<b>2: Improvements to Data Quality</b>				
Patient Demographic Data Quality	<ul style="list-style-type: none"> <li>Assist health care organizations in improving the quality of patient demographic data and in preventing and managing duplicate records</li> </ul>	<ul style="list-style-type: none"> <li>Patient Demographic Data Quality Framework Whitepaper</li> </ul>	<ul style="list-style-type: none"> <li>Patient Demographic Data Quality Framework Pilot</li> </ul>	1, 4
<b>3: Expanded Data Sharing</b>				
Health Data Security Layer	<ul style="list-style-type: none"> <li>Collaboration with OpenID Foundation to create a trusted security layer that enables patients to control sharing of their data</li> </ul>	<ul style="list-style-type: none"> <li>HEART Working Group Specifications</li> </ul>	<ul style="list-style-type: none"> <li>Move Health Data Forward Challenge</li> <li>Proving the Potential: A Health Data and Standards Code-a-Thon</li> </ul>	3

Target Area	Goals	Products / Tools Developed	Relevant Activity / Application	Alignment to ASPE IAA Objective
Tools for the Creation of a Longitudinal Patient Record	<ul style="list-style-type: none"> <li>• Increase accessibility of FHIR to developers with less expertise on health care area</li> <li>• Demonstration of versatility of FHIR-based access to patient data</li> </ul>	<ul style="list-style-type: none"> <li>• FHIR-based Patient Data Toolkit</li> <li>• Claims Data Importer</li> <li>• Record Merge Tool</li> </ul>	<ul style="list-style-type: none"> <li>• Proving the Potential: A Health Data and Standards Code-a-Thon</li> </ul>	3
	<ul style="list-style-type: none"> <li>• Facilitate combination of clinical and claims information related to same individuals</li> </ul>			
Novel Approaches to Health Data Sharing	<ul style="list-style-type: none"> <li>• Exploration of the potential to use distributed ledger technology to health IT challenges; more specifically, to discover new ways to securely share patient data</li> </ul>	<ul style="list-style-type: none"> <li>• 77 Listed submissions for the challenge<sup>52</sup></li> <li>• Challenge and Code-A-Thon artifacts detailing the events</li> </ul>	<ul style="list-style-type: none"> <li>• Challenge: “Use of Blockchain in Health IT and Health-Related Research”</li> <li>• Blockchain in Healthcare Code-A-Thon</li> </ul>	3, 4
<b>4: Data Standardization</b>				
Record Completeness	<ul style="list-style-type: none"> <li>• Assessment of the existence of a core set of data elements in a patient record</li> </ul>	<ul style="list-style-type: none"> <li>• Patient Record Scorecard Application</li> </ul>	<ul style="list-style-type: none"> <li>• Proving the Potential: A Health Data and Standards Code-a-Thon</li> </ul>	1

<sup>52</sup><https://oncprojecttracking.healthit.gov/wiki/display/TechLabI/Blockchain+Challenge+on+ONC+Tech+Lab#BlockchainChallengeonONCTechLab-Bank>

Target Area	Goals	Products / Tools Developed	Relevant Activity / Application	Alignment to ASPE IAA Objective
Health Data Provenance Challenge	<ul style="list-style-type: none"> <li>• Advancement of work in improving availability of information about the data to increase researcher and other data user confidence</li> </ul>	<ul style="list-style-type: none"> <li>• Challenge artifacts detailing the events and lessons learned</li> </ul>	<ul style="list-style-type: none"> <li>• Oh the Places Data Goes Health Data Provenance Challenge</li> </ul>	1
Provider Data	<ul style="list-style-type: none"> <li>• Advancement of accuracy and completeness of provider directory data</li> </ul>	<ul style="list-style-type: none"> <li>• NPPES Open API</li> <li>• OAuth Client for NPPES Open API</li> <li>• Provider data validation tools</li> </ul>	<ul style="list-style-type: none"> <li>• Provider Directory Workshop</li> </ul>	5